



# King Saud University

## Journal of Saudi Chemical Society

[www.ksu.edu.sa](http://www.ksu.edu.sa)  
[www.sciencedirect.com](http://www.sciencedirect.com)



### ORIGINAL ARTICLE

## QSAR based modeling of hepatitis C virus NS5B inhibitors

A.K. Srivastava, Avni Pandey \*, Akanchha Srivastava, Neerja Shukla

*QSAR Research Laboratory, Department of Chemistry, University of Allahabad, Allahabad 211002, India*

Received 28 June 2009; accepted 22 October 2009

Available online 11 November 2010

#### KEYWORDS

QSAR;  
 Physicochemical properties;  
 Regression analysis;  
 HCV NS5B polymerase  
 inhibitors

**Abstract** QSAR analysis on a series of 13 derivatives of 5-hydroxy-3-(2H)-pyridazinone as hepatitis C virus NS5B polymerase inhibitors was performed. The inhibitory activities of these compounds were found to increase with density ( $D$ ), index of refraction ( $I_{or}$ ), surface tension ( $St$ ), electronegativity ( $X_{eq}$ ) and decrease with the balaben centric index ( $B_{ac}$ ), balaben distance connectivity index ( $J$ ) and zero order connectivity ( $X^0$ ) along with the indicator parameter. Excellent results were obtained from regression upon introduction of the indicator parameter. The results are critically discussed on the basis of regression data and cross validation technique.

© 2010 King Saud University. Open access under [CC BY-NC-ND license](#).

### 1. Introduction

Hepatitis C virus (HCV) causes chronic liver diseases. Infected people will develop chronic histological changes in the liver with a high risk of advancing to cirrhosis or hepatocellular carcinoma (Wong and Lee, 2006). The treatment against HCV is a combination therapy of pegylated interferon and ribavirin. Unfortunately, this treatment has 50% benefits and is associated with adverse events. Therefore there is an urgent need for the development of effective HCV therapies.

With this rational in mind, we describe herein the quantitative structure activity relationship (QSAR) studies of a series of HCV NS5B polymerase inhibitors (Zhou et al., 2008), aiming at identifying the physicochemical properties that govern the inhibitory activity of HCV NS5B polymerase inhibitors. It is envisioned that these studies will produce models that can be used for future designing of new analogues with higher potency.

### 2. Results and discussion

QSAR studies were performed on the compounds listed in Table 1, where the biological activity  $IC_{50}$  is a measure of inhibitory activity. All the QSAR reported herein were derived by us and were not reported with the original data set taken from the literature as reference. We have used Hansch analysis for developing these models.

An attempt has been made to correlate the activities of the compounds with physicochemical parameters such as density ( $D$ ) (Acd-Lab software), index of refraction ( $I_{or}$ ) (Acd-Lab software), surface tension ( $St$ ) (Acd-Lab software) and electronegativity ( $X_{eq}$ ) (Acd-Lab software) for a set of HCV NS5B

\* Corresponding author.

E-mail addresses: [draksrivastava1@rediffmail.com](mailto:draksrivastava1@rediffmail.com) (A.K. Srivastava), [avnipandey@rediffmail.com](mailto:avnipandey@rediffmail.com) (A. Pandey), [akanchha\\_au@rediffmail.com](mailto:akanchha_au@rediffmail.com) (A. Srivastava), [neerjashuklaau2007@rediffmail.com](mailto:neerjashuklaau2007@rediffmail.com) (N. Shukla).

1319-6103 © 2010 King Saud University.

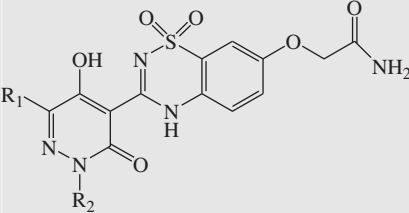



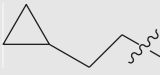





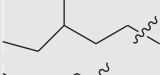

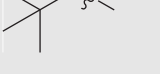



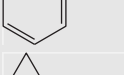








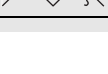
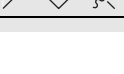
Open access under [CC BY-NC-ND license](#).

doi:10.1016/j.jscs.2010.09.007



Production and hosting by Elsevier

**Table 1** Biological activity and physicochemical data for HCV NS5B polymerase inhibitors.

											
S. No.	$R_1$	$R_2$	D	Ior	St	Bac	$J$	$X^0$	Xeq	$I_1$	pIC <sub>50</sub>
1			1.590	1.730	68.700	77	1.495	25.328	2.423	0	6.469
2			1.720	1.800	81.500	40	1.301	24.742	2.437	0	7.000
3			1.530	1.708	65.200	94	1.502	26.250	2.412	1	6.620
4			1.500	1.697	63.400	94	1.524	26.250	2.403	0	6.004
5			1.550	1.717	66.600	78	1.492	26.035	2.412	0	6.000
6			1.560	1.719	67.100	91	1.518	25.543	2.423	0	6.076
7			1.720	1.800	81.500	40	1.312	24.742	2.437	0	6.268
8			1.620	1.766	75.400	40	1.315	26.156	2.453	0	6.268
9			1.720	1.800	81.500	40	1.301	24.742	2.437	0	6.921
10			1.720	1.800	81.500	40	1.301	24.742	2.437	0	6.854
11			1.650	1.753	73.100	77	1.495	25.328	2.437	0	6.824
12			1.790	1.830	87.400	40	1.301	24.742	2.451	0	6.854
13			1.420	1.649	55.500	128	1.742	25.913	2.379	0	5.292

polymerase inhibitors were calculated by AcD-Lab Chem Sketch software. The balaben centric index (Bac) (DRAGON software), balaben distance connectivity index ( $J$ ) (DRAGON software) and zero order connectivity ( $X^0$ ) (DRAGON software) for a set of HCV NS5B polymerase inhibitors were calculated by DRAGON software.

Indicator parameter is the dummy parameter sometimes used for accounting those structural features not covered in any molecular descriptor used. The details of such parameter, used in the present study are given in the results and discussion section (Table 1). The multiregression analysis used to derive

the correlation was executed with SPSS (7.5) program. Initially, we have used the Pogliani's quality factor  $Q$  for investigating predictive power of the various parameters and finally, the cross validation techniques to prove our findings.

The series of HCV NS5B polymerase inhibitors, their inhibitory activities, physicochemical parameters and indicator parameter are reported in Table 1.

The autocorrelation between the used parameters was checked and is shown in Table 2. The results show that all the physicochemical parameters are orthogonal and can be subjected to multiregression analysis with indicator parameter.

**Table 2** Autocorrelation matrix between physicochemical and indicator parameters.

	$I_1$	$D$	Ior	St	Bac	$J$	$X^0$	Xeq
$I_1$	1.000							
$D$	-0.243	1.000						
Ior	-0.233	0.992	1.000					
St	-0.235	0.992	0.999	1.000				
Bac	0.256	-0.916	-0.954	-0.948	1.000			
$J$	0.140	-0.902	-0.945	-0.939	0.985	1.000		
$X^0$	0.384	-0.842	-0.794	-0.797	0.654	0.612	1.000	
Xeq	-0.192	0.878	0.899	0.892	-0.905	-0.893	-0.557	1.000

The statistically significant models obtained are as follows:

$$\text{pIC}_{50} = 4.544(\pm 1.251)D + 0.622(\pm 0.473)I_1 - 0.955 \quad (1)$$

$$n = 12, R = 0.940, R^2 = 0.884, R_A^2 = 0.858, \text{SE} = 0.194,$$

$$F_{[2,9]} = 34.261, Q = 4.845$$

$$\text{pIC}_{50} = 9.075(\pm 2.946)\text{Ior} + 0.595(\pm 0.544)I_1 - 9.476 \quad (2)$$

$$n = 12, R = 0.920, R^2 = 0.846, R_A^2 = 0.811, \text{SE} = 0.224,$$

$$F_{[2,9]} = 24.654, Q = 4.107$$

$$\text{pIC}_{50} = 0.051(\pm 0.017)\text{St} + 0.597(\pm 0.570)I_1 + 2.699 \quad (3)$$

$$n = 12, R = 0.912, R^2 = 0.831, R_A^2 = 0.794, \text{SE} = 0.234,$$

$$F_{[2,9]} = 22.192, Q = 3.897$$

$$\text{pIC}_{50} = -0.015(\pm 0.007)\text{Bac} + 0.605(\pm 0.732)I_1 + 7.446 \quad (4)$$

$$n = 12, R = 0.851, R^2 = 0.724, R_A^2 = 0.663, \text{SE} = 0.300,$$

$$F_{[2,9]} = 11.814, Q = 2.837$$

$$\text{pIC}_{50} = -3.166(\pm 3.028)J + 0.417(\pm 1.449)I_1 + 10.958 \quad (5)$$

$$n = 12, R = 0.847, R^2 = 0.717, R_A^2 = 0.654, \text{SE} = 0.304,$$

$$F_{[2,9]} = 11.403, Q = 2.786$$

$$\text{pIC}_{50} = -0.719(\pm 0.768)X^0 + 0.809(\pm 1.677)I_1 + 24.692 \quad (6)$$

$$n = 12, R = 0.819, R^2 = 0.670, R_A^2 = 0.597, \text{SE} = 0.328,$$


$$F_{[2,9]} = 9.151, Q = 2.497$$

$$\text{pIC}_{50} = 20.200(\pm 21.062)\text{Xeq} + 0.491(\pm 1.553)I_1 - 42.602 \quad (7)$$

$$n = 12, R = 0.825, R^2 = 0.681, R_A^2 = 0.610, \text{SE} = 0.322,$$

$$F_{[2,9]} = 9.600, Q = 2.562$$

In all the models, the statistical parameter ' $n$ ' is the number of data points, ' $R$ ' is the correlation coefficient, ' $R^2$ ' is the coefficient of determination, ' $R_A^2$ ' is the adjusted  $R^2$ , ' $\text{SE}$ ' is the standard error of estimate, ' $F$ ' is the F ratio (Diudea, 2000; Bikash et al., 2003), ' $Q$ ' is the quality of fit (Pogliani, 1994; Pogliani, 1996), and data within the parenthesis is the confidence interval at 95% level. Compound number 7 is treated as outlier since this compound having a benzyl  $R^2$  substituent displayed a weaker antiviral potency.

From the above regression equations, it is clear that the coefficients of  $D$ , Ior, St and Xeq are positive which indicate that denser, bulkier and more electronegative groups are beneficial for the activity. The negative coefficients of Bac and  $J$  indicate that they have a negative effect on the inhibitory activity. The negative coefficient of  $X^0$  suggests that less zero order branching is favorable. The indicator  $I_1$  for  group at  $R^2$

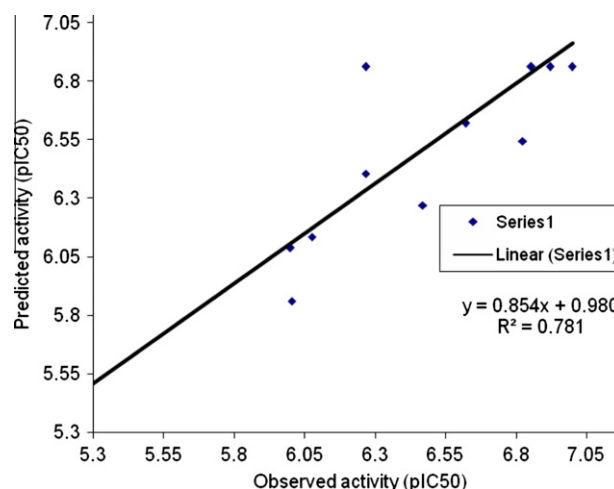
site has positive coefficient indicating thereby this group shows a positive influence in determining the activity. The robustness and applicability of the above QSAR equations were further

confirmed by the cross validation technique like leave one out method (Cramer et al., 1988; Podlgar and Ferguson, 2000).

Predicted and residual values for the derived QSAR models are given in Table 3. Predicted values are the calculated activities of the equation and the residual values are the differences between the observed biological activities and the calculated activities and are found to be low. The calculated  $F$  value is greater than  $F$  theoretical value ( $F_{[2,9]} = 4.26$ ) for all the significant equation. The plot of observed  $\text{pIC}_{50}$  versus predicted

**Table 3** Predicted and residual values for Eq. (1).

S. No.	Observed	Predicted	Residual
1	6.469	6.27	0.198
2	7	6.861	0.139
3	6.62	6.62	0
4	6.004	5.861	0.143
5	6	6.088	-0.088
6	6.076	6.134	-0.058
7	6.268	6.861	-
8	6.268	6.406	-0.139
9	6.921	6.861	0.06
10	6.854	6.861	-0.007
11	6.824	6.543	0.281
12	6.854	7.179	-0.325
13	5.292	5.498	-0.205

**Figure 1** A Plot of comparison between observed and predicted activities for Eq. (1).

**Table 4** Cross validation parameters and predictive error of coefficient of correlation (PE) for the proposed models.

Eq. No.	<i>n</i>	Parameter	PRESS	SSY	PRESS/SSY	$R^2_{cv}$	PSE	<i>R</i>	1- $R^2$	PE	6PE
Eq. (1)	12	$D + I_1$	0.340	2.590	0.131	0.869	0.168	0.940	0.116	0.022	0.134
Eq. (2)	12	$I_{or} + I_1$	0.452	2.478	0.182	0.818	0.194	0.920	0.154	0.030	0.178
Eq. (3)	12	$St + I_1$	0.494	2.436	0.203	0.797	0.203	0.912	0.168	0.032	0.194
Eq. (4)	12	$Bac + I_1$	0.808	2.122	0.381	0.619	0.259	0.851	0.276	0.053	0.318
Eq. (5)	12	$J + I_1$	0.829	2.101	0.395	0.605	0.263	0.847	0.283	0.054	0.324
Eq. (6)	12	$X^0 + I_1$	0.966	1.964	0.492	0.508	0.284	0.819	0.329	0.063	0.378
Eq. (7)	12	$X_{eq} + I_1$	0.935	1.995	0.469	0.531	0.279	0.825	0.319	0.061	0.366

pIC<sub>50</sub> based on the Eq. (1) is shown in graph (Fig. 1) and the predicted  $R^2$  was found to be fairly large.

Where, observed activities are the biological activities IC<sub>50</sub>, which are measures of inhibitory activities, predicted values are the calculated activities of the equations and residual values are the differences between the observed biological activities and the calculated activities.

### 2.1. Cross validation

The cross validation analysis was performed using leave one out (LOO) method (Cramer et al., 1988; Podlgar and Ferguson, 2000), in which one compound is removed from the data set and the activity is correlated using the rest of the data set. The cross-validated  $R^2$  was found to be very close to the value of  $R^2$  for the entire data set and hence these models can be termed as statistically significant.

Cross validation provides the values of PRESS, SSY and  $R^2_{cv}$  and PSE from which we can test the predictive power of the proposed model. The meanings of these cross-validated parameters are given as a footnote to the Table 4. It is argued that PRESS is a good estimate of the real predictive error of the model and if it is smaller than SSY the model predicts better than chance and can be considered statistically significant. Furthermore, the ratio PRESS/SSY can be used to calculate approximate confidence intervals of prediction of new compound. To be a reasonable QSAR model PRESS/SSY should be smaller than 0.4 and for our models the value of this ratio which is smaller than 0.1 indicates that they are excellent models. Also, if PRESS value is transformed in a dimensionless term by relating it to the initial sum of squares, we obtain  $R^2_{cv}$  i.e. the complement to the traces of unexplained variance over the total variance. The PRESS and  $R^2_{cv}$  have good properties. However, for practical purposes of end users the use of square root of PRESS/*n*, which is called predictive square error (PSE), is more directly related to the uncertainty of the predictions. The PSE values also support our results. The calculated cross-validated parameters confirm the validity of the models.

### 2.2. Predictive error of coefficient of correlation (PE)

The predictive error of coefficient of correlation (PE) (Chaterjee et al., 2000) is yet another parameter used to decide the predictive power of the proposed models. We have calculated PE value of all the proposed models and they are reported in Table 4. It is argued that of


- i)  $R < PE$ , then correlation is not significant;
- ii)  $R > PE$ ; several times (at least three times), then correlation is indicated; and if
- iii)  $R > 6PE$ , then the correlation is definitely good.

For all the models developed the condition  $R > 6PE$  and hence they can be said to have good predictive power.

The data presented in Table 4 indicate that in all the cases,  $R$  is much larger than 6PE indicating that all the proposed models have represented high degree of correlation.

## 3. Conclusions

On the basis of the above discussions two general conclusions can be drawn as below.

- i) More denser, more bulkier and more electronegative groups with less branching should be used in the future drug modeling.
- ii) The indicator  $I_1$  for  group at  $R^2$  site is preferentially favorable.

## References

- Accl-Lab software for calculating the referred physicochemical parameters. Chem Sketch. <[www.acclabs.com/acclabs-rss-feed.xml](http://www.acclabs.com/acclabs-rss-feed.xml)>.
- Bikash, D., Shovanlal, G., Subrata, B., Soma, S., Tarun, J., 2003. Bioorg. Med Chem. 11, 5493.
- Chaterjee, S., Hadi, A.S., Price, B., 2000. Regression Analysis by Example, third ed. Wiley VCH, New York.
- Cramer III, R.D., Bunce, J.D., Patterson, D.E., 1988. Quant. Struct. Act. Relat. 7, 18.
- Diudea, M.V. (Ed.), 2000. QSPR/QSAR Studies for Molecular Descriptors, Nova Science, and Hunting don.
- DRAGON software for calculation of topological and connectivity indices. <[www.disal.unimib.it](http://www.disal.unimib.it)>.
- Podlgar, B.L., Ferguson, D.M., 2000. Drug. Des. Discov. 4, 17.
- Pogliani, L., 1994. Amino acids 6, 141.
- Pogliani, L., 1996. J. Phys. Chem. 100, 18065.
- Wong, T., Lee, S.S., 2006. CMAJ 174, 649.
- Zhou, Y., Li, L., Dragovich, P.S., Murphy, D.E., Tran, C.V., Ruebsam, F., Webber, S.E., Shah, A.M., Tsan, M., Averill, A., Showalter, R.E., Patel, R., Han, Q., Zhao, Q., Hermann, T., Kissinger, C.R., LeBrun, L., Sergeeva, M.V., 2008. Bioorg. & Med. Chem. Lett. 18, 1419–1424.